Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population

CSIRO PUBLISHING

Animal Production Science, 2010, **50**, 1004–1010 www.publish.csiro.au/journals/an

Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population

H. D. Daetwyler[A,E], *J. M. Hickey*[B], *J. M. Henshall*[C], *S. Dominik*[C], *B. Gredler*[D], *J. H. J. van der Werf*[B] *and B. J. Hayes*[A]

[A]Biosciences Research Division, Department of Primary Industries, 1 Park Drive, Bundoora, Vic. 3083, Australia.
[B]School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia.
[C]CSIRO Livestock Industries, New England Highway, Armidale, NSW 2350, Australia.
[D]Department of Sustainable Agricultural Systems, University of Natural Resources and Applied Life Sciences Vienna, Gregor Mendel Str. 33, A-1180 Vienna, Austria.
[E]Corresponding author. Email: hans.daetwyler@dpi.vic.gov.au

Abstract. Estimated breeding values for the selection of more profitable sheep for the sheep meat and wool industries are currently based on pedigree and phenotypic records. With the advent of a medium-density DNA marker array, which genotypes ~50 000 ovine single nucleotide polymorphisms, a third source of information has become available. The aim of this paper was to determine whether this genomic information can be used to predict estimated breeding values for wool and meat traits. The effects of all single nucleotide polymorphism markers in a multi-breed sheep reference population of 7180 individuals with phenotypic records were estimated to derive prediction equations for genomic estimated breeding values (GEBV) for greasy fleece weight, fibre diameter, staple strength, breech wrinkle score, weight at ultrasound scanning, scanned eye muscle depth and scanned fat depth. Five hundred and forty industry sires with very accurate Australian sheep breeding values were used as a validation population and the accuracies of GEBV were assessed according to correlations between GEBV and Australian sheep breeding values . The accuracies of GEBV ranged from 0.15 to 0.79 for wool traits in Merino sheep and from –0.07 to 0.57 for meat traits in all breeds studied. Merino industry sires tended to have more accurate GEBV than terminal and maternal breeds because the reference population consisted mainly of Merino haplotypes. The lower accuracy for terminal and maternal breeds suggests that the density of genetic markers used was not high enough for accurate across-breed prediction of marker effects. Our results indicate that an increase in the size of the reference population will increase the accuracy of GEBV.

Additional keywords: genomic selection, single nucleotide polymorphism.

Introduction

Genetic improvement of meat and wool sheep has focussed on traits that are economically important and relatively easy to measure. This has been effective because some important wool traits are highly heritable and are measurable in both sexes (van der Werf 2009). However, measurement of some traits that contribute towards profitability is expensive or difficult, for example, measurement of slaughter traits in meat breeds and fertility traits in wool sheep. The use of genomic data in the form of single nucleotide polymorphisms (SNP) has been proposed as a way to increase the accuracy of selection of such traits and, therefore, their genetic improvement. The advantage of genomic information is that it enables selection for traits for which phenotypes are not available at the time of selection, namely the selection of juveniles for wool, slaughter or fertility traits.

Genomic prediction methods rely on strong associations between SNP and the quantitative trait loci (QTL) that affect the traits of interest (Meuwissen *et al*. 2001). Genomic prediction can potentially capture all of the genetic variation in a trait if the genetic marker map is dense enough. Two main genomic prediction methods are currently used. One is an adaptation of best linear unbiased prediction, in which a genomic relationship matrix is used in place of a pedigree-derived relationship matrix. This method is referred to as GBLUP (NejatiJavaremi *et al*. 1997; Villanueva *et al*. 2005; Hayes *et al*. 2009*c*). GBLUP assumes a normal distribution for SNP effects, which is equivalent to assuming there is a very large number of mutations with very small effects. Other methods for predicting GEBV involve different assumptions: (1) that there are some SNP with a moderate-to-large effect and many SNP with very small effects (BayesA) and (2) that some of the SNP are not associated with mutations affecting the trait and thus have no effect at all (BayesB and BayesSSVS; Meuwissen *et al*. 2001; Verbyla *et al*. 2009; Pong-Wong and Hadjipavlou 2010). These two groups of methods have similar accuracies (correlations between genomic breeding values and traditional breeding values) for most traits (Hayes *et al*. 2009*a*; VanRaden *et al*. 2009) and it has been shown by simulation that this equivalency occurs when many QTL affect a trait (Daetwyler *et al*. 2010).

If the accuracy of genomic estimated breeding values (GEBV) were as high as the square root of the heritability of a trait, which is equivalent to the accuracy of a phenotypic record, genetic gain

 10.1071/AN10096 1836-0939/10/121004

could be increased by up to 40% in sheep breeding programs (van der Werf 2009). The economic impact of genomic prediction of sheep traits depends on whether the increased profitability from the extra genetic gain outweighs the cost of genotyping. Modelling has shown that the net present value of meat sheep production could increase by 25% over the next 25 years if genomic prediction was used to increase the rate of genetic progress by 40% (Banks and van der Werf 2009).

Many specialised breeds are currently used in the Australian sheep industry. Merino sheep are used to produce wool, terminal breeds such as the Polled Dorset and Whiteface Suffolk are favoured by meat producers and maternal breeds such as the Border Leicester and Coopworth are kept for their mothering and reproductive abilities. Although this strategy is useful from an industry standpoint, it complicates genomic prediction because linkage disequilibrium (LD) patterns may not persist across breeds. With multi-breed data, and at current SNP densities, a particular marker allele may not be consistently associated with the same QTL allele in all breeds, making accurate prediction across breeds difficult (De Roos *et al*. 2008). If the phase between SNP and QTL is not consistent between sheep breeds, a large number of phenotype and genotype records would be needed for a particular breed to enable accurate predictions to be made for that breed (De Roos *et al*. 2009; Hayes *et al*. 2009*b*; Ibanz-Escriche *et al*. 2009).

The objective of this study was to evaluate the accuracies of GEBV for meat and wool traits using a large reference population. Two genomic prediction methods were used: GBLUP and BayesA, to investigate their respective genomic prediction accuracy.

Materials and methods

Phenotypic data

The reference population data consisted of multi-breed sheep data from the Sheep Cooperative Research Centre information nucleus flock (INF) and the Sheep Genomics Falkiner Memorial Field Station flock (FMFS) with both phenotypic and genotypic records.

The INF animals were located at eight sites across Australia and the FMFS animals were raised in Deniliquin, New South Wales. Various breeds were represented in both datasets (Table 1). The dams had a strong Merino background in both datasets, whereas the sires were either from maternal, terminal or Merino breeds. While the Merino sheep were mostly purebred, the remaining breeds were mainly represented by their crosses with Merino ewes. Therefore, a significant proportion of the reference population was crossbred with few purebred individuals from terminal and maternal breeds.

The following traits were evaluated: yearling greasy fleece weight (GFW), yearling fibre diameter (FD), yearling staple strength (SS), early breech wrinkle score (EBRWR), late breech wrinkle score (LBRWR), weight at ultrasound scanning (SC_WT), scanned eye muscle depth (SEMD) and scanned fat depth (SFAT). Whereas GFW is a measure of wool growth, FD, the average thickness of individual wool fibres, and SS, the minimum force per density unit required to rupture a staple, are wool quality traits. Breech wrinkle score (BRWR) was defined as the degree of skin wrinkling at the tail set and down the hind legs and was scored on a scale of 1–5, in which 5 is very wrinkly. As high BRWR increases the incidence of flystrike, selective breeding for lower BRWR may reduce flystrike (James 2006). Although most Australian lamb producers are paid according to carcass weight, there is interest in increasing the relative sizes of quality cuts of meat, such as the eye muscle, and in improving carcass leanness by reducing subcutaneous fat depth. Table 2 shows reference population sizes, means, standard deviations (s.d.) and heritabilities for each trait in both resource flocks. Summary statistics were sufficiently similar to justify pooling the INF and FMFS data. Extreme values for fixed effects and traits were removed if they were >4 s.d. from the means for the INF or FMFS.

Genotypic data

All animals were genotyped using the Illumina 50K ovine SNP chip (Illumina Inc., San Diego, CA, USA), which reacts to 54 977 SNP. The following quality control measures were applied to the SNP data: SNP were removed if they had a call rate of <95%, a GC

Table 1. Number of progeny by breed of sire (SBreed) and dam (DBreed) for the Cooperative Research Centre information nucleus flock (before backslash) and the Sheep Genomics Falkiner Research Station flock (after backslash)
BL, Border Leicester; EF, East Friesian; MER, Merino; PD, Polled Dorset; WS, White Suffolk

SBreed	DBreed				
	MER	PD	WS	BL × MER	Total
MER	1305\2077	0\1	0\2	0\0	1305\2080
PD	459\44	0\19	0\32	387\153	846\248
PD × WS	0\61	0\46	0\29	0\205	0\341
BL	790\95	0\31	0\24	0\155	790\305
BL × EF	0\43	0\19	0\19	0\106	0\187
WS	307\39	0\13	0\19	280\109	587\180
Texel	64\0	0\0	0\0	62\0	126\0
Booroola	59\0	0\0	0\0	0\0	59\0
Suffolk	53\0	0\0	0\0	34\0	87\0
Corriedale	4\0	0\0	0\0	0\0	4\0
Coopworth	2\44	0\18	0\15	0\85	2\162
Southdown	1\0	0\0	0\0	0\0	1\0
Total	3044\2403	0\147	0\140	763\813	3807\3541

Table 2. Phenotypic data, number of records (N-ref) and heritabilities h^2 with standard error (s.e.) for eight traits for the reference population of the Sheep Cooperative Research Centre information nucleus flock (INF) and the Sheep Genomics Falkiner Research Station flock (FMFS)

EBRWR, early breech wrinkle score; FD, fibre diameter; GFW, greasy fleece weight; LBRWR, late breech wrinkle score; SC_WT, weight at ultrasound scanning; SEMD, scanned eye muscle depth; SFAT, scanned fat depth; SS, staple strength

Trait	INF			FMFS			N-ref	h^2 ± s.e.
	Mean	s.d.	*n*	Mean	s.d.	*n*		
GFW	3.5	0.94	1800	3.1	0.95	1541	3341	0.55 ± 0.10[A]
FD	17.3	1.73	1304	23.6	6.04	1528	2831	0.75 ± 0.06[A]
SS	33.4	11.1	927	34.1	8.3	1544	2471	0.43 ± 0.07[A]
EBRWR	2.1	1.2	3791	1.2	0.5	3309	7100	0.34 ± 0.07[A]
LBRWR	2.5	1.0	1275	1.3	0.5	1530	2805	–
SC_WT	42.2	8.5	3644	38.6	7.2	3537	7180	0.27 ± 0.04[B]
SEMD	25.6	4.6	3643	21.0	3.7	3523	7166	0.23 ± 0.03[B]
SFAT	2.9	1.3	3642	2.5	0.9	3521	7163	0.22 ± 0.03[B]

[A]S. Hatcher, NSW Department of Primary Industries, unpublished results based on CRC INF data.
[B]Mortimer *et al.* (2010). Based on CRC INF data.

score (proportion of guanine-cytosine pairs) of <0.6, a minor allele frequency of <0.01, a SNP heterozygosity of >3 s.d. from the mean (mean heterozygosity, 0.374; s.d., 0.129), were out of Hardy–Weinberg equilibrium (a *P*-value cut-off of 1^{-15}), had no genome location or were in >0.99 LD with another SNP on the chip. After these measures were applied, 48 640 SNP were used. Data for genotyped animals were removed if their genotype call rate was <0.9, this reduced the number of genotyped animals from 3863 to 3807. Missing genotypes were imputed using fastPHASE (Scheet and Stephens 2006).

Validation population

The prediction equations were tested using a validation population. Genomic prediction accuracy was evaluated as the statistical correlation between GEBV and Australian sheep breeding values (ASBV). This means that the theoretical maximum for the accuracy is the accuracy of the ASBV. One can transform accuracies based on statistical correlation alone by dividing them by the accuracy of the ASBV (Verbyla *et al.* 2009), which will make the theoretical maximum of this new accuracy measure one. Unless otherwise stated, we report and discuss the accuracy based on the statistical correlation of GEBV and ASBV.

Validation rams needed to have ASBV with accuracy greater than 0.5 and 540 rams distributed across several breeds satisfied these criteria. ASBV for these industry rams were calculated without using the phenotypic data from the INF flock. Information about ASBV definitions can be found at the following website maintained by Australian Wool Innovation Ltd and Meat and Livestock Australia (AWI and MLA 2010). Table 3 lists the number of animals available and their average ASBV accuracies for each trait. One hundred and twelve of the validation sires had progeny in the INF. Sires in the INF were chosen to maximise the connectedness with the Australian sheep flock by sampling artificial insemination sires from as many studs as possible. The validation rams also originated in the general Australian sheep population, thus the relationship between the reference and validation population is expected to be moderate.

Accuracies were calculated within breeds: within the Merino breed, accuracies were calculated for maternal (mainly Border

Table 3. Mean accuracy of Australian sheep breeding values in validation rams per trait in Merino (MER), Merino superfine (MER-SF), Merino fine (MER-F), Merino strong (MER-S), maternal breed Border Leicester (BL), terminal breed (TERM), terminal Polled Dorset (TERM-PD), terminal White Suffolk (TERM-WS) and number of validation sires available for each breed or Merino wool type

BRWR, breech wrinkle score; FD, fibre diameter; GFW, greasy fleece weight; SC_WT, weight at ultrasound scanning; SEMD, scanned eye muscle depth; SFAT, scanned fat depth; SS, staple strength

Trait	MER	MER-SF	MER-F	MER-S	BL	TERM	TERM-PD	TERM-WS
Number	187	19	96	70	57	218	108	99
Yearling GFW	0.91	0.91	0.92	0.91	–	–	–	–
Adult GFW	0.87	0.91	0.87	0.87	–	–	–	–
Yearling FD	0.95	0.94	0.95	0.94	–	–	–	–
Adult FD	0.92	0.94	0.92	0.91	–	–	–	–
Yearling SS	0.84	0.83	0.84	0.83	–	–	–	–
BRWR	0.79	0.73	0.83	0.74	–	–	–	–
SC_WT	0.91	0.89	0.91	0.93	0.91	0.93	0.94	0.93
SEMD	0.83	0.75	0.82	0.86	0.89	0.92	0.94	0.93
SFAT	0.79	0.73	0.78	0.82	0.90	0.91	0.93	0.91

Leicester and some Coopworth) and terminal breeds (Polled Dorset and White Suffolk) and for superfine-, fine- and strong-wool Merino types. The Merino wool types are a subset of the Merino rams based on FD.

Statistical models

The following fixed effects were fitted in all trait models: sire breed, dam breed (Merino or Border Leicester × Merino for the INF and proportion of Merino, Border Leicester, Polled Dorset and White Suffolk for the FMFS), sex, birth type, rearing type, contemporary group (birth year × site × management group) and age-at-trait recording. In the FMFS, full dam information was unavailable for one of the lambings and dam breed proportions were inferred from maternal haplotypes using the program structure (Pritchard *et al*. 2000). In addition, SC_WT was fitted in the analysis of SEMD and SFAT.

For GBLUP, the following model involving a realised relationship matrix based on markers was fitted to the phenotypes using ASReml (Gilmour *et al*. 2000):

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{e}$$

where **y** is a vector of phenotypic records, **X** is an design matrix relating the fixed effects (as described above) to the animal, **b** is a vector of fixed effects, **Z** is a design matrix relating animal effects to phenotypes, **g** is a vector of additive genetic effects and **e** is the vector of residuals. The following distributions were assumed: $\mathbf{g} \sim N(0, \sigma_g^2\mathbf{G})$ and $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$. **G** was calculated as in (Hayes *et al*. 2009*c*).

The model for BayesA was similar to that used for GBLUP but instead of calculating a relationship based on markers, each marker was fitted individually in vector **v** and **Z** connected records and SNP effects:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zv} + \mathbf{Wu} + \mathbf{e}$$

where a polygenic effect was fitted with design matrix **W** relating polygenic effects to records and the vector of additive polygenic effects, **u**. Each SNP had its own variance, $\mathbf{v}_i \sim N(0, \sigma_{v_i}^2)$ and polygenic effects distributed as $\mathbf{u} \sim N(0, \sigma_u^2\mathbf{A})$, where **A** is the pedigree-derived relationship matrix. Pedigree was traced back up to five generations, including dam pedigree at some INF sites. The prior for $\sigma_{v_i}^2$ was an inverse Chi-square distribution (degrees of freedom, 4.012) (Meuwissen *et al*. 2001). For each trait, 10 parallel runs, each with a Markov Chain Monte Carlo chain of 50 000 iterations and a burn-in of 10 000 iterations, were conducted. The results of the 10 chains were averaged to calculate GEBV.

Results and discussion

The accuracy of GEBV for wool traits was only evaluated for Merino sheep because there were a limited number of observations for the other breed types and the ASBV accuracy for wool traits of maternal and terminal validation rams was low or zero in some cases. In Merino sheep, the correlations between yearling GEBV and yearling ASBV for wool traits ranged from 0.23 to 0.79 for GBLUP and from 0.15 to 0.74 for BayesA (Table 4). Both GFW and FD had accuracies in excess of 0.70. However, the accuracy for SS was less than 0.70, which may be partly explained by two factors: the number of SS records was low and SS has a lower heritability than either GFW or FD (Table 2). These accuracy trends were similar to those for the correlation between yearling GEBV and adult ASBV (Table 5), which, to some extent, is expected as genetic correlations between yearling and adult wool traits are generally >0.60 (Huisman and Brown 2009). However, it is noteworthy because accurate prediction of adult wool ASBV from yearling GEBV facilitates accurate selection of young rams for the adult GFW and FD of their progeny. The accuracy within wool types was generally somewhat lower and more variable than the accuracy for all Merino sheep combined. The between-Merino wool type variation in accuracy was likely increased by some wool types having only a few validation rams available. For example, there were only 19 superfine Merino rams available for validation. The lower correlations between wool types compared with those for the full Merino dataset likely also reflects the fact that some of the SNP predict between-strain differences.

The definition of an appropriate reference set for EBRWR and LBRWR data was difficult because age at recording had differed between the INF and FMFS (INF EBRWR = 35 days, FMFS EBRWR = 243 days, INF LBRWR = 360 days, FMFS LBRWR = 534 days). Therefore, six different ways of combining the data were investigated and evaluated using GBLUP (Table 6). In two scenarios, INF EBRWR and LBRWR alone were evaluated and in four scenarios, all possible combinations were analysed. The main conclusions were as follows. First, LBRWR observations were more predictive of adult BRWR ASBV (higher accuracies were achieved with fewer records for LBRWR than for EBRWR). Second, when accuracies based on the LBRWR of the INF alone (0.40) were compared with those for combinations of INF LBRWR with either FMFS EBRWR (0.48) or FMFS LBRWR

Table 4. Accuracies (correlations between yearling genomic estimated breeding values and yearling Australian sheep breeding values) for wool traits in Merino sheep according to the GBLUP and BayesA methods

FD, fibre diameter; GFW, greasy fleece weight; MER, Merino; MER-F, Merino fine; MER-SF, Merino superfine; MER-S, Merino strong; *n*, number of validation sires; N-ref, size of the reference population; SS, staple strength

Breed	*n*	GFW		FD		SS	
		GBLUP	BayesA	GBLUP	BayesA	GBLUP	BayesA
MER	187	0.73	0.70	0.79	0.74	0.23	0.15
MER-SF	19	0.65	0.58	0.65	0.51	–0.01	0.06
MER-F	96	0.63	0.64	0.57	0.47	0.43	0.34
MER-S	69	0.51	0.40	0.49	0.23	0.18	0.20
N-ref	–	3341	3341	2831	2831	2471	2471

Table 5. Accuracies (correlations of yearling genomic estimated breeding values and adult Australian sheep breeding values) for wool traits in Merino sheep with GBLUP and BayesA

BRWR, breech wrinkle score; E-FMFS, early Falkiner Research Station flock BRWR observations; FD, fibre diameter; GFW, greasy fleece weight; L-INF, late information nucleus BRWR observations; MER, Merino; MER-F, Merino fine; MER-SF, Merino superfine; MER-S, Merino strong; *n*, number of validation sires; N-ref, size of the reference population

Breed	*n*	GFW		FD		BRWR (L-INF + E-FMFS)	
		GBLUP	BayesA	GBLUP	BayesA	GBLUP	BayesA
MER	187	0.71	0.68	0.79	0.77	0.48	0.47
MER-SF	19	0.55	0.47	0.44	0.42	0.27	0.32
MER-F	96	0.61	0.62	0.61	0.55	0.33	0.30
MER-S	69	0.51	0.38	0.43	0.26	0.48	0.48
N-ref	–	3341	3341	2831	2831	4584	4584

Table 6. Accuracy of GBLUP from different combinations of breech wrinkle score (BRWR) data

E, early BRWR observations; FMFS, Sheep Genomics Falkiner Research Station flock; INF, information nucleus flock; L, late BRWR observations; MER, Merino; MER-F, Merino fine; MER-S, Merino strong; MER-SF, Merino superfine; *n*, number of validation sires; N-ref, size of the reference population

Breed	*n*	BRWR					
		E-INF	E-INF + E-FMFS	E-INF + L-FMFS	L-INF	L-INF + E-FMFS	L-INF + L-FMFS
MER	123	0.32	0.41	0.39	0.40	0.48	0.45
MER-SF	10	0.40	0.48	0.47	0.27	0.27	0.32
MER-F	69	0.24	0.28	0.26	0.24	0.33	0.28
MER-S	39	0.28	0.42	0.40	0.37	0.48	0.45
N-ref	–	3791	7100	5317	1275	4584	2801

(0.45), it was evident that combining INF and FMFS data increases genomic prediction accuracy because it increases the number of observations. This trend was also observed for other traits (results not shown). BayesA analysis of BRWR was performed only for the combination scenario yielding the highest accuracy in Table 6 and the results are shown in Table 5.

The analysis of meat traits resulted in accuracies lower than those for wool traits. Accuracies ranged from –0.08 to 0.50 with GBLUP and from –0.04 to 0.57 with BayesA (Table 7). The lower accuracy of meat traits compared with that of wool traits may be related to heritability, which was typically lower for meat traits (Table 2). Transformed accuracies for the meat traits are shown in Table 8 and are higher than statistical correlations, reflecting the fact that ASBV of validation rams had an accuracy of less than one. Overall, Merino sheep achieved higher accuracies than maternal and terminal breeds. This is likely a reflection of the high proportion of Merino haplotypes in the two reference flocks. For SEMD and SFAT, the GEBV accuracy achieved with GBLUP ranged from 0.21 to 0.50. Further increases in accuracy are expected when more of the INF animals are genotyped and the size of the reference population increases. Uncharacteristically low accuracies were observed for the SC_WT of terminal breeds and further study is needed to investigate the causes for this. In contrast, GEBV accuracies for the Merino breed were similar across the three meat traits.

Comparison of results from the GBLUP and BayesA methods showed that GBLUP tended to give a slightly higher accuracy than BayesA for most traits. One reason for the small difference

Table 7. Accuracies for meat traits in Merino, maternal and terminal breeds with GBLUP and BayesA

BL, maternal breed Border Leicester; MER, Merino; MER-F, Merino fine; MER-SF, Merino superfine; MER-S, Merino strong; *n*, number of validation sires; N-ref, size of the reference population; SC_WT, weight at ultrasound scanning; SEMD, scanned eye muscle depth; SFAT, scanned fat depth; TERM, terminal breed; TERM-PD, terminal Polled Dorset; TERM-WS, terminal White Suffolk

Breed	*n*	SC_WT		SEMD		SFAT	
		GBLUP	BayesA	GBLUP	BayesA	GBLUP	BayesA
MER	164	0.49	0.57	0.49	0.39	0.45	0.42
MER-SF	15	0.41	0.55	–0.08	0.12	0.21	–0.04
MER-F	80	0.27	0.48	0.44	0.46	0.49	0.48
MER-S	66	0.22	0.14	0.43	0.30	0.27	0.39
BL	56	0.36	0.24	0.21	0.42	0.12	0.20
TERM	218	0.07	–0.04	0.43	0.19	0.28	0.10
TERM-PD	108	0.17	0.08	0.5	0.34	0.21	0.06
TERM-WS	99	–0.04	–0.07	0.27	0.11	0.27	0.05
N-ref	–	7180	7180	7166	7166	7163	7163

Table 8. Transformed accuracies [cor(GEBV, ASBV)/acc(ASBV)] for meat traits in Merino, maternal and terminal breeds with GBLUP and BayesA

BL, maternal breed Border Leicester; MER, Merino; *n*, number of validation sires; N-ref, size of the reference population; SC_WT, weight at ultrasound scanning; SEMD, scanned eye muscle depth; SFAT, scanned fat depth; TERM, terminal breed; TERM-PD, terminal Polled Dorset; TERM-WS, terminal White Suffolk

Breed	*n*	SC_WT		SEMD		SFAT	
		GBLUP	BayesA	GBLUP	BayesA	GBLUP	BayesA
MER	164	0.54	0.62	0.66	0.47	0.57	0.53
BL	56	0.39	0.27	0.24	0.48	0.13	0.22
TERM	218	0.08	–0.04	0.47	0.20	0.30	0.11
TERM-PD	108	0.18	0.08	0.53	0.36	0.22	0.06
TERM-WS	99	–0.04	–0.08	0.30	0.12	0.29	0.06
N-ref	–	7180	7180	7166	7166	7163	7163

between the methods may be because a polygenic effect was fitted with BayesA but not with GBLUP. Realised accuracies matched analytical expectations well (Goddard 2009; Daetwyler *et al*. 2010). Furthermore, BayesA did not identify many SNP with moderate allele substitution effects. The largest effect was seen in SC_WT (0.1 s.d. units) and there also seem to be a small group of intermediate SNP effects present in this trait which could partially explain why BayesA performed slightly better in Merino sheep SC_WT. However, overall the SNP effects were very small, which indicates that the traits are controlled by a very large number of QTL. In such genetic architectures, GBLUP and Bayesian models will result in similar accuracies, as pointed out earlier.

The general difference in accuracy between the Merino and the terminal breeds indicates that the current SNP marker density is not sufficient for marker effects to be transferable across breeds. This is probably because SNP alleles are not consistently associated with the same QTL alleles in different breeds. Two strategies could be used to achieve higher accuracies. First, the number of genotyped animals from each breed in the reference population could be increased to enable accurate estimation of marker effects for that breed alone. Second, the marker density of the genotyping platform could be increased to increase the level of LD in the marker set. This strategy is used with other species to increase GEBV accuracy. However, in the short term, increases in accuracy for terminal and maternal breeds will likely have to be derived via an increase in the size of the reference populations, because a denser than 50K SNP chip or re-sequence data is currently not available in sheep.

Conclusions

Moderate-to-high GEBV accuracies were achieved when the reference set for estimating marker effects was sufficiently large. Merino sheep tended to have higher accuracies than terminal and maternal breeds because the reference population had a strong Merino background. Given the low numbers of reference animals for the terminal breeds, GEBV accuracies for SEMD and SFAT are encouraging. The current SNP marker density is not high enough to enable prediction of marker effects across breeds. GEBV accuracy was increased by combining the INF and FMFS datasets and further increases in sample size would improve GEBV accuracy.

Acknowledgements

The authors acknowledge the Cooperative Research Centre for Sheep Industry Innovation, and the Sheep Genomics Project, including Meat and Livestock Australia and Australian Wool Innovation Limited, for funding. We thank Daniel Brown and Andrew Swan for providing validation ram ASBV excluding INF data, Klint Gore for managing the phenotypic INF database, Cedric Gondro for performing part of the genotype quality control, as well as Sue Mortimer and Sue Hatcher for providing us with heritability estimates from INF data. Two anonymous reviewers are thanked for their constructive comments, which have improved the manuscript.

References

AWI, MLA (2010) Gain from genetics. Australian Wool and Innovation, Meat and Livestock Australia. Available at http://www.makingmorefromsheep.com.au/gain-from-genetics/index.htm [Verified 20 September 2010]

Banks RG, van der Werf JHJ (2009) Economic evaluation of whole genome selection, using meat sheep as a case study. In 'Proceedings of the 18th Conference of the Association for the Advancement of Animal Breeding and Genetics'. Adelaide, Australia. pp. 430–433. (Association for the Advancement of Animal Breeding and Genetics)

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. *Genetics* **185**, 1021–1031. doi:10.1534/genetics.110.116855

De Roos APW, Hayes BJ, Spelman RJ, Goddard ME (2008) Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* **179**, 1503–1512. doi:10.1534/genetics.107.084301

De Roos APW, Hayes BJ, Goddard ME (2009) Reliability of genomic breeding values across multiple populations. *Genetics* **183**, 1545–1553. doi:10.1534/genetics.109.104935

Gilmour AR, Thompson R, Cullis BR, Wellham SJ (2000) 'ASReml reference manual.' (NSW Department for Primary Industries: Orange, NSW)

Goddard ME (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245–257. doi:10.1007/s10709-008-9308-0

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009*a*) Invited review: genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* **92**, 433–443. doi:10.3168/jds.2008-1646

Hayes BJ, Bowman PJ, Chamberlain AJ, Verbyla KL, Goddard ME (2009*b*) Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics, Selection, Evolution* **41**, 51. doi:10.1186/1297-9686-41-51

Hayes BJ, Visscher PM, Goddard ME (2009*c*) Increased accuracy of artificial selection by using the realized relationship matrix. *Genetical Research* **91**, 47–60. doi:10.1017/S0016672308009981

Huisman AE, Brown DJ (2009) Genetic parameters for bodyweight, wool, and disease resistance and reproduction traits in Merino sheep. 4. Genetic relationships between and within wool traits. *Animal Production Science* **49**, 289–296. doi:10.1071/EA08173

Ibanz-Escriche N, Fernando RL, Toosi A, Dekkers JCM (2009) Genomic selection of purebreds for crossbred performance. *Genetics, Selection, Evolution* **41**, 12. doi:10.1186/1297-9686-41-12

James PJ (2006) Genetic alternatives to mulesing and tail docking in sheep: a review. *Australian Journal of Experimental Agriculture* **46**, 1–18. doi:10.1071/EA05100

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829.

Mortimer SI, van der Werf JHJ, Jacob RH, Pethick DW, Pearce KL, Warner RD, Geesink GH, Hocking Edwards JE, Gardner GE, Ponnampalam EN, Kitessa SM, Ball AJ, Hopkins DL (2010) Preliminary estimates of genetic parameters for carcass and meat quality traits in Australian sheep. *Animal Production Science* **50**, 1135–1144. doi:10.1071/AN10126

NejatiJavaremi A, Smith C, Gibson JP (1997) Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of Animal Science* **75**, 1738–1745.

Pong-Wong R, Hadjipavlou G (2010) A two-step approach combining the Gompertz growth model with genomic selection for longitudinal data. *BMC Proceedings* **4**(Suppl. 1), S4. doi:10.1186/1753-6561-4-s1-s4

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics* **78**, 629–644. doi:10.1086/502802

van der Werf JHJ (2009) Potential benefit of genomic selection in sheep. In 'Proceedings of the 18th Conference of the Association for the Advancement of Animal Breeding and Genetics, Adelaide, Australia'. pp. 38–41. (Association for the Advancement of Animal Breeding and Genetics)

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* **92**, 16–24. doi:10.3168/jds.2008-1514

Verbyla KL, Hayes BJ, Bowman PJ, Goddard ME (2009) Technical note: accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. *Genetics Research* **91**, 307–311. doi:10.1017/S0016672309990243

Villanueva B, Pong-Wong R, Fernandez J, Toro MA (2005) Benefits from marker-assisted selection under an additive polygenic genetic model. *Journal of Animal Science* **83**, 1747–1752.

Manuscript received 17 June 2010, accepted 6 October 2010